ROBUST RUMOR DETECTION ON SOCIAL MEDIA WITH DYNAMIC CONTRASTIVE LEARNING

Mei Guo[1,2] *Chen Chen*[1,2*] *Chunyan Hou*[3] *Yike Wu*[4] *Xiaojie Yuan*[1,2]

[1] College of Computer Science, Nankai University, Tianjin, China
[2] MoE Key Lab of DISSec, Nankai University, Tianjin, China
[3] School of CSE, Tianjin University of Technology, Tianjin, China
[4] School of Information and Communication, Nankai University, Tianjin, China

ABSTRACT

Detecting rumors on social media has become increasingly desirable due to its rapid spread and threat to society and individuals. To capture the temporal patterns of propagation graph, Graph Neural Networks (GNNs) based on dynamic propagation graph are proposed to detect rumors on social media. Those methods face a robustness challenge because each training sample usually contains only a partial propagation structure which may include missing and unreliable interactions in the real scenario. Thus, it is important to improve the robustness of dynamic propagation graph-based models. To address this issue, we propose a novel end-to-end Dynamic Contrastive Learning (DCL) model, which integrates dynamic augmented graphs with contrastive learning. Specifically, GNNs are adopted to encode dynamic propagation graphs and dynamic augmented graphs, and the contrastive loss encourages DCL to learn robust and dynamic representations of posts. Extensive experiments are conducted on two public real-world datasets. Experimental results show that our DCL model significantly improves the effectiveness and robustness compared with the state-of-the-art baselines.

Index Terms— Rumor Detection, Dynamic Graph, Contrastive Learning

1. INTRODUCTION

Social media has become the primary platform for people to access news, express opinions, and engage in social interactions. It also has the risk of the proliferation and rapid spread of rumors and poses serious threats to social stability and economic order. Therefore, it is urgent to detect rumors effectively on social media.

Traditional rumor detection methods rely on manually crafted feature engineering to extract rumor features [1]. Recent deep learning-based rumor detection methods have been proposed, mainly categorized into text-based [2], social context-based [3, 4], and propagation-based approaches [5, 6]. Because Graph Neural Networks (GNNs) have advantages for graph data, some studies have applied GNNs to explore the structure of propagation graph for rumor detection on social media [7, 4]. These GNN models only consider the static propagation structure of the final state of rumors, and ignore the temporal pattern of news propagation graphs. In fact, the spread of social media posts evolves and the dynamics of posts exhibit the temporal propagation pattern. Therefore, GNNs based on dynamic propagation graphs are proposed to model the temporal dynamics and detect rumors [8, 9]. Fig. 1 illustrates an example of dynamic propagation graph which consists of a series of stages.

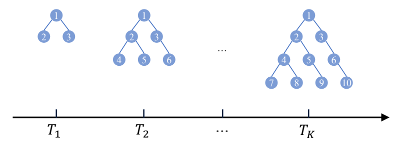


Fig. 1: An example of dynamic propagation graph. Nodes represent posts. Edges represent relationships between nodes. The dynamic propagation process is divided into K stages.

However, these dynamic propagation methods only focus on learning dynamic representations of the propagation graphs and ignore the robustness of rumor detection models. In the real scenario, training samples typically exhibit partial propagation structures that involve missing or unreliable interactions [10]. This fact contributes to the propagation uncertainty and makes it challenging to discover temporal propagation pattern on dynamic propagation graphs. We focus on model robustness, referring to reliable performance under propagation uncertainty or graph sparsity in the early stage.

In this paper, we propose a novel end-to-end Dynamic Contrastive Learning (DCL) model, which integrates the temporal dynamics of propagation graphs with contrastive learning for rumor detection. Specifically, we first divide the propagation process of social media posts into a series of consecutive stages in chronological order by their release time. The graph of each stage represents a static snapshot of propaga-

*Corresponding author.

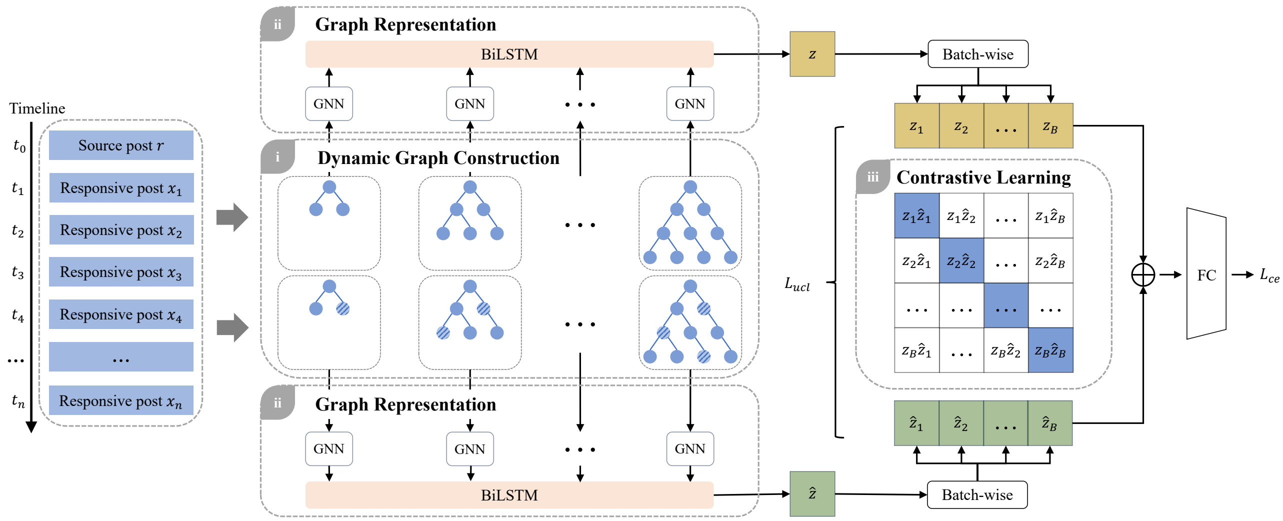


Fig. 2: The framework of the proposed DCL model.

tion graphs, which consists of the source post and response posts until the end of current stage. Then we use graph augmentation methods to construct the corresponding augmented graph for each stage and obtain a series of graph snapshots of dynamic propagation graphs and dynamic augmented graphs. Finally, we train an end-to-end model which integrates contrastive learning to enhance the robustness and effectiveness.

Our contributions are summarized as follows: (1) We propose a novel end-to-end Dynamic Contrastive Learning (DCL) model, which integrates dynamic augmented graphs with contrastive learning to improve the robustness and effectiveness of rumor detection. (2) We study a novel problem of the robustness of GNNs based on dynamic propagation graphs, and explore graph augmentation methods. (3) We conduct extensive experiments on two real-world datasets to demonstrate the effectiveness and robustness of our model.

2. METHODOLOGY

Rumor detection can be defined as a binary classification task. Given an event $c = \{r, x_1, \ldots, x_{n-1}, G\}$, r is the source post, x_j represents the j-th response post. All posts are ordered chronologically and the sequence of release time is denoted as $s = \{t_0, t_1, \ldots, t_{n-1}\}$, where t_j is the release time of the j-th post. $G = \langle V, E\rangle$ is the propagation graph, where V represents posts and E is their relationships. The feature matrix is denoted as $\mathbf{X}$. The adjacency matrix is represented as A. The architecture of DCL is shown in Fig. 2.

2.1. Dynamic Graph Construction

Dynamic Propagation Graph Construction. Some works have proposed to capture the dynamic pattern of the propagation graph [8]. We extend these works and attempt three dynamic graph construction methods: time snapshots, sequential snapshots, and fixed-time snapshots. We divide the propagation graph of event c into K graph snapshots along the release timeline. Each graph snapshot has an equal time interval $\Delta t = \frac{t_{n-1}-t_0}{K}$. Then we construct a series of graph snapshots of dynamic propagation graph $G = \{G_1, G_2, \ldots, G_K\}$, where $G_k = \langle V_k, E_k\rangle$ is the k-th graph snapshot. $V_k = \{c_k | s_k \leq k\Delta t\}$ is the node set. E_k is the edge set. The adjacency matrix is A_k. The feature matrix is denotes as $\mathbf{X}_k$.

Dynamic Augmented Graph Construction. Propagation-based rumor detection models usually suffer from missing and unreliable interactions. Graph augmentation can alleviate this issue. We explore three graph augmentation strategies: node mask, edge drop, and feature mask. Given a graph snapshot G_k of dynamic propagation graph, we randomly mask some nodes to obtain the augmented graph snapshot $\hat{G}_k$. The feature matrix is represented as $\hat{\mathbf{X}}_k = \mathbf{M} \odot \mathbf{X}_k$, where $\mathbf{M} \in \{0,1\}^{n_k \times d}$ contains $n_k \times r_{mask}$ rows of zero vectors. r_{mask} is the masking rate. $\odot$ denotes the element-wise multiplication. Then we obtain a series of graph snapshots of dynamic augmented graph $\hat{G} = \{\hat{G}_1, \hat{G}_2, \ldots, \hat{G}_K\}$, where $\hat{G}_k = \langle V_k, E_k\rangle$ is the k-th augmented graph snapshot. The adjacency matrix is A_k and the feature matrix is $\hat{\mathbf{X}}_k$.

2.2. Graph Representation

We use Graph Isomorphism Network (GIN) [11] to capture the representations of the graph snapshots of dynamic propagation graph $G = \{G_1, G_2, \ldots, G_K\}$ and the graph snapshots of dynamic augmented graph $\hat{G} = \{\hat{G}_1, \hat{G}_2, \ldots, \hat{G}_K\}$. We employ a two-layer GIN to obtain the representations H_k of G_k. Then the mean-pooling operator is used to aggregate the information to obtain graph representations $\mathbf{h}_k$.

$$H_k = \mathrm{GIN}(A_k, \mathbf{X}_k), \mathbf{h}_k = \mathrm{MEAN}(H_k) \tag{1}$$

Finally, the dynamic propagation graph is represented as $\mathbf{h} = \{\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_K\}$. Similarly, the dynamic augmented graph $\hat{\mathbf{h}} = \{\hat{\mathbf{h}}_1, \hat{\mathbf{h}}_2, \ldots, \hat{\mathbf{h}}_K\}$ is obtained in the same manner.

Then we adopt Bidirectional Long Short-Term Memory (BiLSTM) [12] to model graph representation sequences in both forward and backward directions. We concatenate the forward state and the backward state to obtain the graph representation z encoded by BiLSTM, where CONCAT represents the concatenation operation.

$$\overrightarrow{z} = \overrightarrow{\text{LSTM}}(\mathbf{h}), \overleftarrow{z} = \overleftarrow{\text{LSTM}}(\mathbf{h}) \tag{2}$$

$$z = \text{CONCAT}(\overrightarrow{z}, \overleftarrow{z}) \tag{3}$$

Finally, we obtain the representation of dynamic propagation graph $Z = \{z_1, z_2, \ldots, z_{\mathcal{B}}\}$. The representation of dynamic augmented graph $\hat{Z} = \{\hat{z}_1, \hat{z}_2, \ldots, \hat{z}_{\mathcal{B}}\}$ is obtained in the same manner. $\mathcal{B}$ denotes the batch size.

2.3. Contrastive Learning

We formulate an unsupervised contrastive learning method to maximize the consistency between the graph representations of the dynamic propagation graph and its corresponding dynamic augmented graph. Specifically, given Z and $\hat{Z}$, we construct $\mathcal{B}^2$ representation pairs $\{(z_i, \hat{z}_j)|i, j = 1, \ldots, \mathcal{B}\}$. Among them, only the $\mathcal{B}$ pairs with $i = j$ are correct matches, which are defined as positive samples. The remaining $\mathcal{B}^2 - \mathcal{B}$ pairs are wrong matches and are regarded as negative samples. Our contrastive learning aims to maximize the similarity of the $\mathcal{B}$ correct matches while minimizing the similarity of the $\mathcal{B}^2 - \mathcal{B}$ wrong matches. We construct the similarity matrix $\Psi \in \mathbb{R}^{\mathcal{B} \times \mathcal{B}}$ to capture the pairwise similarities and apply the multi-class N-pair loss [13] by considering the row-wise cross entropy loss with respect to the row index.

$$\Psi = \exp(\frac{\text{sim}(Z, \hat{Z})}{\tau}) \tag{4}$$

$$\mathcal{L}_{ucl} = \text{CE}(\Psi, \text{I}) \tag{5}$$

where $\text{sim}(\cdot)$ denotes the cosine similarity $\text{sim}(Z, \hat{Z}) = \frac{Z\hat{Z}^{\text{T}}}{||Z|| \cdot ||\hat{Z}||}$. τ is a temperature parameter. $\text{I} \in \mathbb{R}^{\mathcal{B} \times \mathcal{B}}$ represents the identity matrix, and CE denotes the cross entropy loss.

2.4. Training Objective

We concatenate Z and $\hat{Z}$ to integrate the information as $\text{Z} = \text{CONCAT}(Z, \hat{Z})$. Then Z is fed into fully connected layers and a softmax layer, and the output is calculated as

$$\hat{y} = \text{softmax}(W_f \text{Z} + b_f) \tag{6}$$

where $\hat{y} \in \mathbb{R}^{\mathcal{B} \times \mathcal{C}}$ is the predicted probability distribution. W_f and b_f are weight matrix and bias vector, respectively.

$$\mathcal{L}_{ce} = -\frac{1}{\mathcal{B}} \sum_{b=1}^{\mathcal{B}} \sum_{c=1}^{\mathcal{C}} y_{b,c} \log(\hat{y}_{b,c}) \tag{7}$$

	# Events	# Posts	# Non-Rumors	# Rumors	# Avg. time length
RumorEval	245	4145	112	133	12 Hours
TWITTER	1077	60207	564	513	416 Hours

Table 1: Statistics of the datasets. "#" is "the number of".

Method	Class	RumorEval				TWITTER			
		Acc.	Prec.	Rec.	F1	Acc.	Prec.	Rec.	F1
Bi-GCN	F	0.7115	0.7314	0.7599	0.7454	0.7591	0.7595	0.7419	0.7506
	T		0.7042	0.6562	0.6794		0.7771	0.7732	0.7751
EBGCN	F	0.6952	0.7100	0.7652	0.7366	0.7539	0.7449	0.7428	0.7438
	T		0.6812	0.6125	0.6450		0.7657	0.7640	0.7648
GACL	F	0.7250	0.7974	0.7385	0.7668	0.7609	0.7987	0.6781	0.7335
	T		0.7591	0.7091	0.7332		0.7454	0.8377	0.7889
RDEA	F	0.7321	0.7513	0.7890	0.7697	0.7855	0.7942	0.7496	0.7713
	T		0.7434	0.6648	0.7019		0.7857	0.8181	0.8016
TrustRD	F	0.7267	0.7298	0.8175	0.7712	0.7695	0.7751	0.7359	0.7550
	T		0.7617	0.6195	0.6833		0.7749	0.7974	0.7860
DynGCN	F	0.7377	0.7471	0.7617	0.7543	0.7693	0.7647	0.7495	0.7570
	T		0.7092	0.6823	0.6955		0.7759	0.7893	0.7825
DCL	F	**0.8121**	**0.8413**	**0.8215**	**0.8313**	**0.8308**	**0.8368**	**0.8034**	**0.8198**
	T		**0.7991**	**0.8009**	**0.8000**		**0.8319**	**0.8554**	**0.8435**

Table 2: Rumor detection results on RumorEval and TWITTER datasets. Abbrev.: Rumor (F), Non-Rumor (T).

where $y_{b,c}$ denotes ground-truth label and $\hat{y}_{b,c} \in \mathbb{R}^{1 \times \mathcal{C}}$ denotes the predicted probability distribution that index $b \in \{1, \ldots, \mathcal{B}\}$ belongs to class $c \in \{1, \ldots, \mathcal{C}\}$.

Finally, our proposed end-to-end model can be trained by minimizing the unsupervised contrastive loss $\mathcal{L}_{ucl}$ and the cross-entropy loss $\mathcal{L}_{ce}$:

$$\mathcal{L} = \mathcal{L}_{ce} + \lambda \mathcal{L}_{ucl} \tag{8}$$

where λ is the trade-off parameter.

3. EXPERIMENTS

3.1. Experimental Setup

Dataset. We conduct experiments on two public real-world datasets: RumorEval [14] and TWITTER [15], which are collected from Twitter. Both datasets contain post texts, timestamps, and propagation information. The datasets have two labels: Rumor (F) and Non-Rumor (T). We retain events with more than three comments in each dataset for our experiments. The statistics of the two datasets are shown in Table 1.

Comparison Models. We compare our proposed model against several strong baselines. Bi-GCN [7] and EBGCN [16] focus on bi-directional propagation. GACL [17], RDEA [18], and TrustRD utilize contrastive learning. DynGCN [8] incorporates temporal information.

Implementation Details. We conduct 5-fold cross-validation. The node hidden dimension is 128. The masking rate is 0.2. The temperature parameter τ is set to 0.8 for RumorEval and 0.4 for TWITTER. The trade-off parameter λ is set to 0.1. The number of graph snapshots K is set to 3. The evaluation metrics used are Accuracy (Acc.), Precision (Prec.), Recall (Rec.), and F1-score (F1).

Model	RumorEval		TWITTER	
	Acc.	F1	Acc.	F1
DCL	**0.8121**	**0.8313**	**0.8308**	**0.8198**
w/o Prop	0.8017	0.8181	0.8223	0.8086
w/o Aug	0.8054	0.8272	0.8190	0.8034
w/o CL	0.8050	0.8229	0.8250	0.8091
w/o Dyn	0.7850	0.8071	0.8198	0.8100
sequential	0.8042	0.8218	0.8245	0.8095
fixed-time	0.8075	0.8265	0.8219	0.8057
edge drop	0.8025	0.8176	0.8244	0.8128
feature mask	0.8004	0.8196	0.8092	0.7899

Table 3: Results of ablation study on two datasets.

3.2. Overall Performance

Our DCL model outperforms all baseline models across evaluation metrics on RumorEval and TWITTER datasets, as shown in Table 2. BiGCN and EBGCN have the worst performance. These models only capture spatial information, which makes them vulnerable to noise. GACL, RDEA, and TrustRD integrate contrastive learning, achieving better results than those based solely on propagation structures. DynGCN captures both temporal and spatial information. However, DynGCN fails to effectively learn robust representations of dynamic propagation graphs. In contrast, our model emphasizes the temporal dynamics of propagation graphs and learns robust representations through contrastive learning. By using dynamic propagation graphs and dynamic augmented graphs simultaneously, our model enhances the ability to capture both spatial and temporal information. Additionally, contrastive learning enables our model to obtain high-quality feature representations and achieve superior performance.

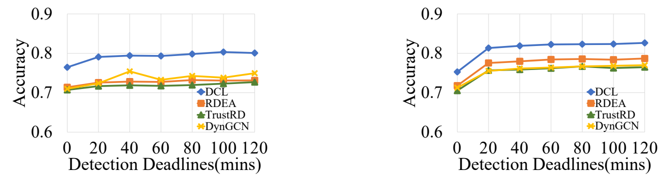
3.3. Ablation Study

As shown in Table 3, we conduct an ablation study to analyze the effect of each DCL module. “w/o Prop” removes dynamic propagation graphs, “w/o Aug” deletes dynamic augmented graphs, “w/o CL” neglects contrastive learning loss, and “w/o Dyn” uses static graphs instead of dynamic graphs. The results validate that spatial structure information, the augmented strategy, contrastive learning and temporal information collectively enable the model to achieve superior performance. We can observe that time snapshots used in DCL effectively capture the temporal evolution patterns, and the node mask used in DCL achieves the best performance.

3.4. Robustness Study

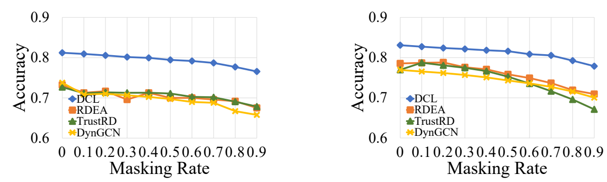
Early Rumor Detection. Fig. 3 shows that at time 0, all models exhibit poor performance due to limited training data. Subsequently, all models show improvement in accuracy as the detection deadlines elapse. DCL consistently achieves higher accuracy than other models at each deadline and has the best robustness in early rumor detection.

Incremental Masking Rates. The increase in masking rate leads to a gradual decline in accuracy for all models in Fig. 4, which reflects the robustness of the models. The accuracy of DCL decreases more slowly and consistently outperforms the comparison models, which demonstrates that DCL is robust and can effectively alleviate the effect of missing data.



(a) RumorEval dataset (b) TWITTER dataset

Fig. 3: Results of early rumor detection on two datasets.



(a) RumorEval dataset (b) TWITTER dataset

Fig. 4: Results of incremental masking rates on two datasets.

4. CONCLUSION

In this paper, we propose a novel end-to-end Dynamic Contrastive Learning (DCL) model for rumor detection. Dynamic propagation graphs and dynamic augmented graphs are fed into our proposed model, where a Graph Isomorphism Network (GIN) and a BiLSTM are used to capture their representations. Moreover, we train the model by incorporating contrastive learning on dynamic augmented graphs. Experiments on two public datasets demonstrate that the DCL model outperforms state-of-the-art baselines in both robustness and effectiveness.

5. ACKNOWLEDGMENTS

This work was partially supported by the National Natural Science Foundation of China (No. 62302245, 62372252, 62172237, 62077031, 62176028), Ministry of Education of the People’s Republic of China Humanities and Social Sciences Youth Foundation (No. 23YJCZH240), and the NSFC-General Technology Joint Fund for Basic Research (No. U1936105, U1936206).

6. REFERENCES

[1] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete, "Information credibility on twitter," in *Proceedings of the 20th international conference on World wide web*, 2011, pp. 675–684.

[2] Yaqian Dun, Kefei Tu, Chen Chen, Chunyan Hou, and Xiaojie Yuan, "Kan: Knowledge-aware attention network for fake news detection," in *Proceedings of the AAAI conference on artificial intelligence*, 2021, vol. 35, pp. 81–89.

[3] Lin Tian, Xiuzhen Jenny Zhang, and Jey Han Lau, "Duck: Rumour detection on social media by modelling user and comment propagation networks," in *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2022, pp. 4939–4949.

[4] Erxue Min, Yu Rong, Yatao Bian, Tingyang Xu, Peilin Zhao, Junzhou Huang, and Sophia Ananiadou, "Divide-and-conquer: Post-user interaction network for fake news detection on social media," in *Proceedings of the ACM web conference 2022*, 2022, pp. 1148–1158.

[5] Jing Ma, Wei Gao, and Kam-Fai Wong, "Rumor detection on twitter with tree-structured recursive neural networks," Association for Computational Linguistics, 2018.

[6] Guanghui Ma, Chunming Hu, Ling Ge, Junfan Chen, Hong Zhang, and Richong Zhang, "Towards robust false information detection on social networks with contrastive learning," in *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, 2022, pp. 1441–1450.

[7] Tian Bian, Xi Xiao, Tingyang Xu, Peilin Zhao, Wenbing Huang, Yu Rong, and Junzhou Huang, "Rumor detection on social media with bi-directional graph convolutional networks," in *Proceedings of the AAAI conference on artificial intelligence*, 2020, vol. 34, pp. 549–556.

[8] Jiho Choi, Taewook Ko, Younhyuk Choi, Hyungho Byun, and Chong-kwon Kim, "Dynamic graph convolutional networks with attention mechanism for rumor detection on social media," *Plos one*, vol. 16, no. 8, pp. e0256039, 2021.

[9] Shuai Xu, Jianqiu Xu, Shuo Yu, and Bohan Li, "Identifying disinformation from online social media via dynamic modeling across propagation stages," in *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, 2024, pp. 2712–2721.

[10] Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Kai-Cheng Yang, Alessandro Flammini, and Filippo Menczer, "The spread of low-credibility content by social bots," *Nature communications*, vol. 9, no. 1, pp. 1–9, 2018.

[11] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka, "How powerful are graph neural networks?," in *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. 2019, OpenReview.net.

[12] Sepp Hochreiter and Jürgen Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[13] Kihyuk Sohn, "Improved deep metric learning with multi-class n-pair loss objective," *Advances in neural information processing systems*, vol. 29, 2016.

[14] Leon Derczynski, Kalina Bontcheva, Maria Liakata, Rob Procter, Geraldine Wong Sak Hoi, and Arkaitz Zubiaga, "SemEval-2017 task 8: RumourEval: Determining rumour veracity and support for rumours," in *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, Vancouver, Canada, Aug. 2017, pp. 69–76, Association for Computational Linguistics.

[15] Hongzhan Lin, Jing Ma, Liangliang Chen, Zhiwei Yang, Mingfei Cheng, and Chen Guang, "Detect rumors in microblog posts for low-resource domains via adversarial contrastive learning," in *Findings of the Association for Computational Linguistics: NAACL 2022*, 2022, pp. 2543–2556.

[16] Lingwei Wei, Dou Hu, Wei Zhou, Zhaojuan Yue, and Songlin Hu, "Towards propagation uncertainty: Edge-enhanced Bayesian graph convolutional networks for rumor detection," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Online, Aug. 2021, pp. 3845–3854, Association for Computational Linguistics.

[17] Tiening Sun, Zhong Qian, Sujun Dong, Peifeng Li, and Qiaoming Zhu, "Rumor detection on social media with graph adversarial contrastive learning," in *Proceedings of the ACM Web Conference 2022*, 2022, pp. 2789–2797.

[18] Zhenyu He, Ce Li, Fan Zhou, and Yi Yang, "Rumor detection on social media with event augmentations," in *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*, 2021, pp. 2020–2024.